

ABSTRACT OF THE DISCLOSURE

Methods of document expansion for a speech retrieval document by a recognizer.

A database of vectors of automatic transcriptions of documents is accessed and the vectors are truncated by removing all terms that are not recognizable by the recognizer to create truncated vectors. Terms in the vectors are then weighted to associate the truncated vectors with the untruncated vectors. Terms not recognized by the recognizer are then added back to the weighted, truncated vectors. The retrieval effectiveness may then be measured.